BEST AVAILABLE IMAGES

Defective images within this document are accurate representation of The original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

In vivo and *in vitro* analyses of an intron-encoded DNA endonuclease from yeast mitochondria. Recognition site by site-directed mutagenesis

B.Sargueil, D.Hatat, A.Delahodde and C.Jacq
Laboratoire de Génétique Moléculaire, CNRS UA 1302, Ecole Normale Supérieure, 46 rue d'Ulm, 75230 Paris Cedex 05, France

Received July 27, 1990; Revised and Accepted September 3, 1990

ABSTRACT

The pal 4 nuclease (termed *I-Sce II*) is encoded in the group I al 4 intron of the COX I gene of *Saccharomyces cerevisiae*. It introduces a specific double-strand break at the junction of the two exons A4-A5 and thus mediates the insertion of the intron into an intronless strain. To define the sequence recognized by pal 4 we introduced 35 single mutations in its target sequence and examined their cleavage properties either *in vivo* in *E. coli* (when different forms of the pal 4 proteins were artificially produced) or *in vitro* with mitochondrial extracts of a mutant yeast strain blocked in the splicing of the al 4 intron. We also detected the pal 4 DNA endonuclease activity in extracts of the wild type strain. The results suggest that 6 to 9 noncontiguous bases in the 17 base-pair region examined are necessary for pal 4 nuclease to bind and cleave its recognition site. We observed that the pal 4 nuclease specificity can be significantly different with the different forms of the protein thus explaining why only some forms are highly toxic in *E. coli*. This study shows that pal 4 recognition site is a complex phenomenon and this might have evolutionary implications on the transfer properties of the intron.

INTRODUCTION

The majority of group I and group II introns in organelle genes contain an intron open reading frame (ORF) which is, most of the time, in phase with the upstream exons, and which codes for proteins involved in recombinational processes of RNA and DNA sequences. These different proteins are proposed to act as RNA maturases (1), DNA endonucleases (2, 3), DNA recombinases (4), and as reverse transcriptases (5). The different activities are mostly poorly understood. Concerning group I intronic proteins, very good evidence of their RNA maturase or DNA endonuclease activities have been presented (6, 7, 8, 9, 10). These two activities, which play a key role in two concerted events of yeast-mitochondrial-genome structure and expression, are especially interesting to study.

DNA endonucleases (8, 9) are, at least in the case of group I introns, responsible for intron insertion by specifically introducing a double strand cut in the vicinity of the exon junction. In the two better documented cases, the mitochondrial DNA endonucleases recognize an 18 base-pair sequence overlapping the two flanking exons (2, 11, 12). In contrast, a prokaryotic intron-encoded endonuclease from a mobile group I intron (13) seems to have a larger recognition sequence and it cleaves at a distance (up to 24 bp) from the insertion site (14).

RNA maturases (4, 7) are required for the splicing of one or two introns. RNA maturases are found in three out of the eight group I intron ORF in the yeast strain 777-3A. None of these maturase-containing introns have been shown to be mobile (15) and at least in the case of the pbI 4 RNA maturase no DNA endonuclease activity has been found (11). This suggests that, at least in the wild-type cell, the two activities might be exclusive and carried out by distinct proteins.

Interestingly, yeast mitochondrial DNA endonucleases and RNA maturases belong to the same family of group I (16) intron-encoded proteins. In fact, two closely related intron-encoded proteins happens each to carry one of the two type of activities. They are coded in the bI 4 intron of cytochrome b gene and in the al 4 intron of COXI gene from the *Saccharomyces cerevisiae* strain 777-3A. pbI 4 is required for the *in vivo* splicing of the two introns bI 4 and al 4 and has thus a pleiotropic RNA maturase activity (17, 18, 19); the pal 4 protein has a specific DNA endonuclease activity and it promotes the insertion of the DNA intron into the genome of an intronless strain (12, 11). We have shown, by making universal genetic code equivalents of the mitochondrial genes, that the RNA maturase activity of pbI 4 is present in a 250 amino acid long protein that can be cytoplasmically translated and targeted to the yeast mitochondria. The DNA endonuclease activity can be observed with a 254 amino acid long pal 4 protein expressed in *E. coli* (11). In the rest of this work we will refer to the two short intronic products mentioned above as the RNA maturase pbI 4 and as the DNA endonuclease pal 4 . However, it must be kept in mind that, *in vivo*, these two proteins are translated as long chimeric proteins encompassing the upstream exons. If these precursors are certainly processed in the splicing mutants, nothing is known about the wild type situation because of the scarcity of those proteins (20).

In this study, we have extended the analysis of the *in vivo* and *in vitro* site recognition properties of the pal 4 protein. We have taken advantage of the engineered pal 4 coding sequence

expressed in *E. coli* to analyze, with a double plasmid system, the properties of mutated forms of the recognition sequence. Surprisingly, many single mutations within the 19 bp region examined do not affect cleavage by pal 4. These recognition properties are however slightly different between variants of the *E. coli*-made protein. We have also compared their *in vivo* properties to the *in vitro* properties of a crude mitochondrial extract of a mutant accumulating the pal 4 protein. These studies, which examine the recognition properties of an intronic DNA endonuclease that acts upon an exonic coding sequence, seem to indicate some relationships between the codon structure and the DNA endonuclease specificity which might have evolutionary implications. Altogether this study shows that pal 4 site recognition is a complex phenomenon similar to the one described for the HO nuclease (21).

MATERIALS AND METHODS

Strains and plasmids

E. coli strain GY7511 (gal K, ilv, his, [λ cI 857ts, N7 N53, Δ Bam ΔH1]) (22), which produces cI857, was used for the *in vivo* assays. RZ 1032 (HfrKL16 PO/45 lysA(6162), dut1, ung1, thi1, relA1, Zdb-279: :Tn10 supE44) (23) and HB2154 (K12, Δ(lac-pro), SupE, thi/F' proA+B+, lacI^q, lacZ DM15, mut1: :Tn10) (24) were used for oligonucleotide-directed mutagenesis.

Yeast strains used for mitochondrial extracts were: CW02 (α,ade2-1, trp1-1, leu2-3, 112 his3-11, 15 ura3 CanR, (intronless cytochrome b gene constructed by M. Labouesse) (18); 777-3A (α, ade1, op1, (rho+)) (25); CO66 (constructed and kindly given by P. Netter), which is isonuclear with 777-3A and is a double mitochondrial mutant (box2-G2590 and 'box9 like' G192) (26); KM 91 is a diploid from the cross of 777-3A and KL14-4A/60/1 (a, His1, trp2 (rho 0)).

Plasmids pUC A4-A5 and pGP al 4 (figure 1) were described in Delahodde et al. (11).

Oligonucleotide mediated mutagenesis

The principle of this method has been described by Kunkel et al. (23). It was used to mutate the ACT codon at position 145 to a CTT codon. The deoxyuridine-substituted template was produced in the RZ1032 strain and isolated following the CTAB procedure (27). Annealing of the phosphorylated oligonucleotide (5' GATATAGATTACTTAATAAAACTGG 3') to the single-stranded template, the elongation reactions and the ligation reactions were according Sambrook et al. (28). The reaction products were introduced into the HB 2154 strain and seven clones randomly chosen were sequenced.

Construction of randomly mutagenized A4-A5 sites

Oligonucleotides were synthesized on an automated DNA synthesizer (Applied Biosystems 381 A). They contain the A4-A5 target sequence plus terminal sequences to allow their insertion into the EcoRI site of the pUC19. Four rounds of synthesis were made in which each base (A, G, T, C) was successively individually degenerated. In order to have an average of one base substitution per molecule, the mixture of phosphoramidites was as following: for A degeneracy 80%A, 7% of each other, for G degeneracy 70% G, 10% of each other, for C degeneracy 65% C, 12% of each other, for T degeneracy 70%T, 10% of each other. Those percentages were calculated with respect to the total amount of each nucleotide in the sequence. Five mutants which could not be obtained by this random approach were completely synthesized.

DNA sequencing

After mutagenesis DNA sequence was checked using chain-termination sequencing (29) according the sequenase protocol (USBC).

Endonuclease assays in *E. coli*

GY 7511 cotransformant clones containing pGPal 4 and pUC19 A4-A5 were grown to stationnary phase at 28°C. Denaturation of cI857 and al 4 production was induced by incubating the cells for 2 hours at 42°C (37° or 40°C has the same effect). Plasmid DNA was then isolated as described in Delahodde et al. (11). Plasmids were digested with SspI and analyzed on a 0.9% agarose gel. For Southern blot analysis, Hybond N+ membrane (Amersham) was used to bind DNA that was transferred with a Vacugene apparatus (LKB). The probe (pUC19 A4A5) was ^{32}P labeled by random priming with the multiprime-DNA-labeling system (Amersham). Hybridization was performed as described by Church et al. (30).

Mitochondrial extract preparation and endonuclease assay

Cells were grown in YEPGal (1% yeast extract, 1% peptone, 2% galactose) and harvested during exponential growth. Mitochondria were isolated as described by Huspedth et al. (31) and further purified on a sucrose discontinuous gradient (20%, 36% and 50% sucrose). Mitochondria were collected, lysed and the endonuclease assay was performed according to Wenzlau et al. (12). The mit+ strain extracts (results shown in Fig. 4) were dialyzed against Tris-HCl 50mM pH 7,5; NaCl 100mM; EDTA 2 mM; DTT 2mM; PMSF 0.1mM and 10% glycerol before the DNA endonuclease assay.

Pulsed-field electrophoresis

E. coli genomic DNA was prepared by the method of Caron (F. Caron personal communication) which is as follow . Cells (10 ml) were harvested during exponential growth phase, washed twice in TE buffer and suspended in 1 ml of buffer A (10% sucrose, 50mM Tris-HCl, pH 8) with 4 mg of lysozyme. The cell suspension was then mixed with 2 ml of low melting agarose. These agarose plugs were incubated in 0.5M EDTA, pH 9, 1% SDS, 1% sarcosyl, and 1 mg/ml proteinase K overnight at 50°C. Before restriction, plugs were washed twice with TE buffer and then once with the incubation buffer. Restriction digests with NotI were done at 37°C overnight.

Pulsed-field electrophoreses were performed according to the CHEF (Contour-clamped Homogenous Electric field) principle (32). We used the LKB 2015 pulsaphor electrophoresis unit and the electrophoretic conditions were the followings: 1.4% agarose gel in 0.25× TBE buffer (22.5 mM Tris-borate, 0.5 mM EDTA). 180 volts during 38 hours with a switching of 30 seconds at 10°C.

RESULTS

In vivo and *in vitro* cleavage of different mutated A4-A5 recognition sites. Activity of the different forms of the *E. coli* synthetized pal 4 DNA endonuclease and activity of crude mitochondrial extracts

It was previously suggested (11, 12) that the recognition site of the pal 4 DNA endonuclease was in an 18 bp long sequence overlapping the A4-A5 splice junction. We constructed 35 single

mutations within the A4-A5 recognition site and we introduced them into a pUC vector to analyze, wether they are recognized and cleaved when the pal 4 protein is produced in the same cell (see figure 1). By using this *in vivo* assay, three different forms of the pal 4 protein were examined. pal 4/A (258) corresponds to the translation, in *E. coli*, of the 3' terminal end (258 amino acids) of the mitochondrial intron open reading frame in which the three codons UGA, AUA and CUN have been changed into their equivalents in the universal genetic code i.e. UGG, AUG and ACN, respectively (33). This pal 4/A protein has already been shown to be toxic to the bacteria if its gene is expressed and can only be stably maintained in *E. coli* in the presence of tightly blocked promoters. In the course of the construction of this gene (11) we observed that a much less toxic form of the protein is obtained if the CUU codon (amino acid 145 when numbered from the first amino acid of the intronic ORF) is not corrected, that is if the protein contains a Leu at position 145 instead of a Thr as suggested by the mitochondrial genetic code. We thus made pal 4/B which is a T to L (145) mutated form of the pal 4/A. We also introduced a second mutation (E to K,122) in a shorter form (249 amino acids) of the pal 4/B protein and this leads to the protein pal 4/C. This last mutation was chosen because it reveals, *in vivo*, an RNA maturase-like property (34) which might alter the DNA endonucleolytic activity.

These three forms of the pal 4 coding sequence were expressed in the presence of different mutated forms of the A4-A5 target sequence and the cleavage efficiencies were compared to that of the wild-type sequence. Examples of such analyses are given in Figure 2. In most of the cases where the target site is cleaved, two stoichiometric homogeneous fragments appear at 2150 bp and 570 bp. This reveals the specific DNA endonuclease activity and indicates that all these DNA ends are fairly stable in *E. coli*.

Such an analysis has also been conducted, *in vitro*, on all 35 mutated target sites with crude mitochondrial extracts of a strain in which RNA splicing of the intron al 4 is blocked because of the absence of the intron bI 4 (strain CWO2 constructed by M. Labouesse; (35)). These *in vitro* results can be compared with the previous *in vivo* approach where cleavage of the wild-type target sequence is similar. It appears that some mutations behave similarly in regard to the four different conditions (for instance G+6) whereas some positions seem to be distinguished by the different forms of the DNA endonuclease activity (for instance T+5). Systematic analysis leads us to four specificity patterns related to the three forms of the pal 4 protein made in *E. coli* and to the mitochondrial extract. These results schematically distinguish three states in which the recognition site is cut like the wild type sequence (long bars), the recognition site is partially cut (middle bars) and the recognition site cut cannot be detected

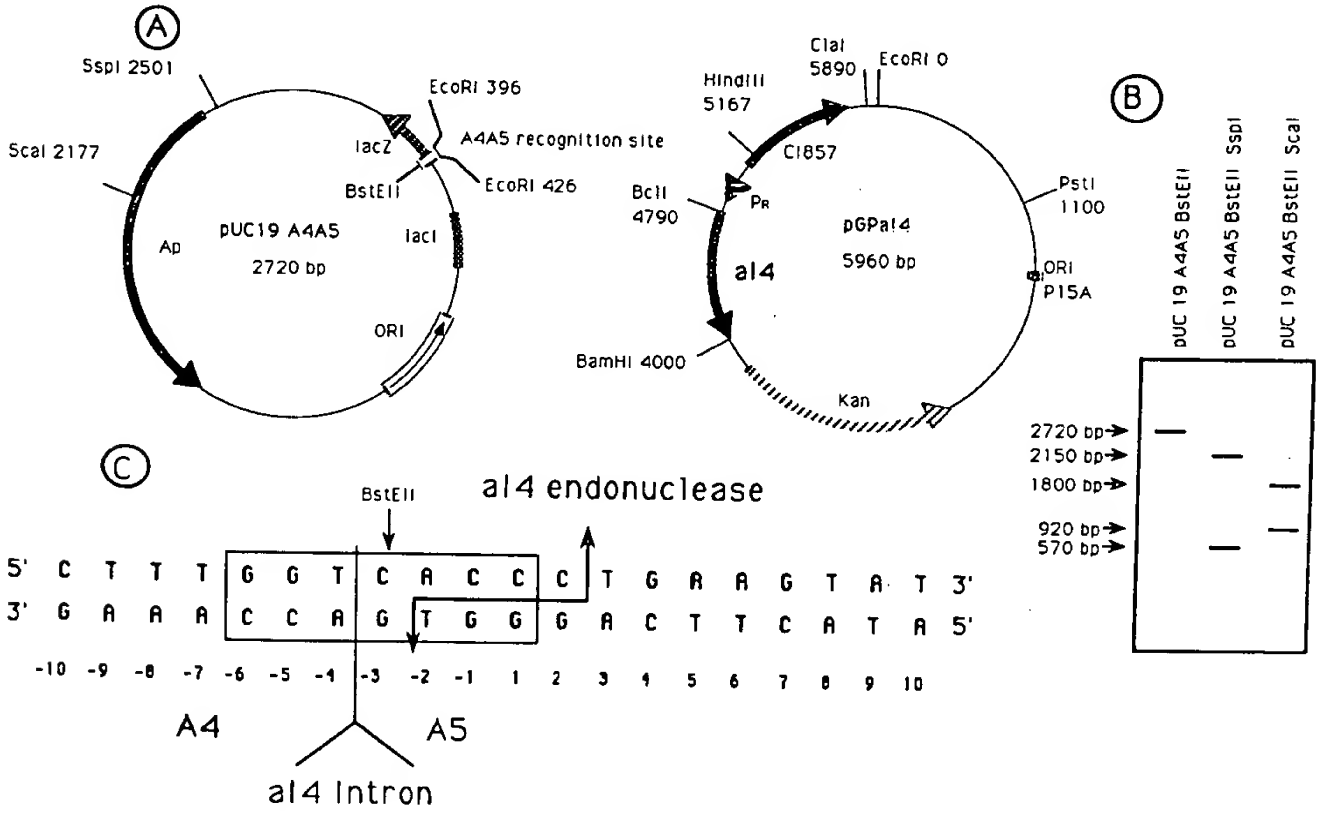


Figure 1. Schematic approach to study, in vivo, the DNA endonucleolytic activity of the pal 4 intronic protein (*I-SceII*). Two compatible plasmids (A),one of them containing either the gene coding for the pal 4 DNA endonuclease, which was adapted to the 'universal' genetic code, and the other one the target sequence A4-A5, were cotransformed into the same *E. coli* strain. The pal 4 coding sequence was under the transcriptional control of the lambda P_R promoter. It was repressed by the thermosensitive CI 857 repressor, which was translated from both the same plasmid and strain GY7511. The mutated A4-A5 recognition sequence (C) was made from synthetic oligodeoxynucleotides and inserted into the EcoRI site of the pUC 19 polylinker sequence. The endonucleolytic activity of the *E. coli*-made pal 4 protein was estimated from Southern blot analyses (B) of the extracted DNA. Accurate characterization of the cleavage site was ensured by taking advantage of the unique BstE II site in the vicinity of the pal 4 cut site (C).

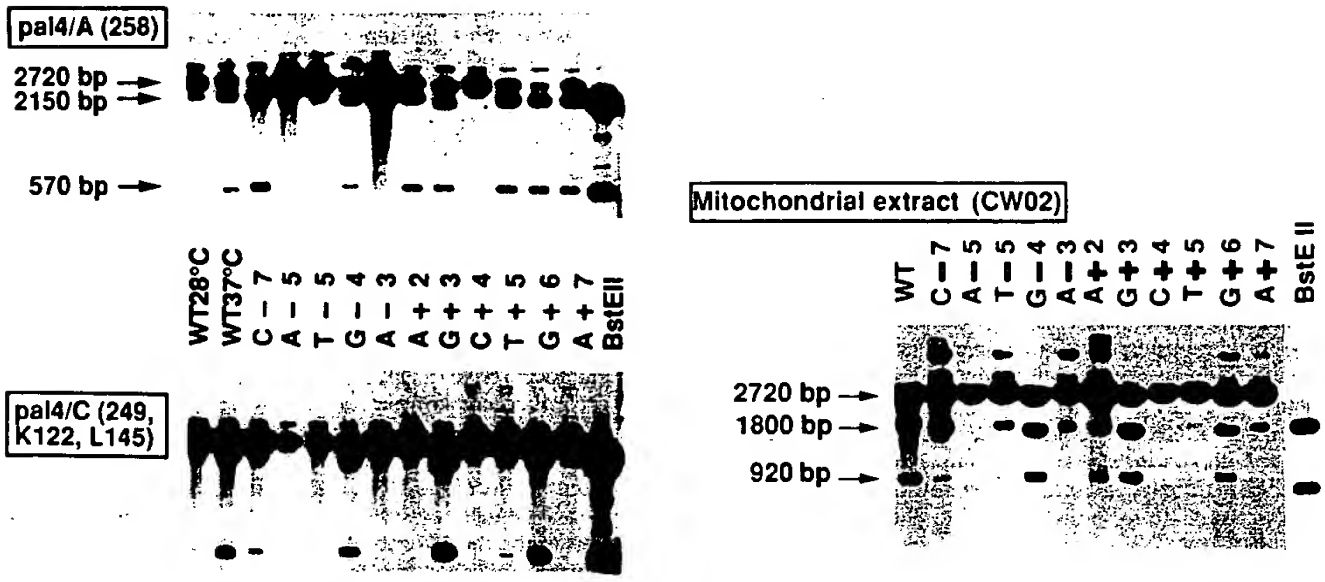


Figure 2. A Southern blot analysis of DNA extracted from *E. coli* transformed with the two plasmids pUC A4-A5 and pGpal 4 (see legend of Fig.1). 35 single mutants in the 19 bp long target sequence were examined either in vivo (left) or in vitro (right) for their ability to be cleaved by the *E. coli* -made pal 4 protein (left) or by a mitochondrial extract of a yeast strain blocked in al4 intron splicing (right). Results of only 11 mutants, named by their position relatively to the cleavage cut (see Fig.1), are shown with the 258 amino acid long protein (pal 4/A 258) or its shorter mutated form (pal 4/C see text); the extracted DNA was cut with Ssp I and hybridized with a pUC 19 probe. The same approach was used in the in vitro analysis (right) except that the DNA was cut with Sca I. The amount of total protein estimated by the BCA protein assay (Pierce Chemical Company) was 16 μg/assay.

(small white bars). Considering the *E. coli* forms of the protein, it is clear that pal 4/A is the least specific in that only three positions (−5; −2 and +4) are critical. At the opposite, pal 4/C tolerates a few variable positions (about four out of eighteen positions) and at least nine positions are stringent positions where none of the examined base changes was tolerated. pal 4/B behaves intermediately and, interestingly, exhibits a pattern somewhat similar to that of the mitochondrial extract (Figure 3,B and D).

It should be stressed that the three types of *in vivo* experiments (A,B,C) described above have been carried under rigorously identical conditions (see Materials and Methods). Only the level of the pal 4 protein could, a priori, vary in the cell due to either differential production or stability. This, however, is unlikely since we observed an equivalent activity of the different forms on the wild-type recognition sequence (Figure 2), and gel electrophoresis analyses of the different transformants did not show significant differences in the level of pal 4 (data not shown).

The yeast pal 4 DNA endonuclease activity can be observed in different mitochondrial genetic contexts

Wenzlau et al. (12) observed in lysates of mitochondria from a mutated strain an al 4-encoded, sequence-specific, DNA endonuclease activity which they did not detect in the corresponding wild-type strain. Following a similar protocol, we looked for the presence of the same specific DNA endonuclease activity in four different genetic contexts (Figure 4). As expected, the strain CW02, which is devoid of the bI 4 RNA maturase and thus cannot splice the al 4 intron(therefore accumulating protein products of that intron), exhibits the presence of an important DNA endonuclease activity. The same specific activity is also present in the double mutant C066, where two homologous intronic proteins, the bI 4 RNA maturase and the al 4 DNA endonuclease, accumulate. More surprisingly, we could observe (Figure 4) the al 4-specific DNA endonuclease activity in the mitochondrial extracts of two wild-type strains: the haploid 777-3A and the diploid KM91. This last observation is all the more interesting since we know that the al 4 intronic translation product responsible for the DNA endonuclease activity is certainly, like the bI 4 translation product, in very low amounts in the wild type cell (20). This clearly suggests that the mitochondrial protein has a very high specific activity or that the mitochondria contain an enhancer of the DNA endonuclease activity; the absence of this enhancer or the presence of an inhibitor in the strain examined by Wenzlau et al. (12) might explain the discrepancy between the two studies. It is difficult to analyze quantitatively the results of Figure 4 since we cannot estimate the concentration of the DNA endonuclease which has not yet been characterized (see discussion). We can only make a rough estimation of the endonuclease activity by considering the total amount of protein used in the different assays. Clearly, when the bI 4 RNA maturase accumulates in the same cell, it does not affect the DNA endonuclease activity (strain C066) and the level of DNA endonuclease activity in the wild type cell is lower than that of the mutant cell (CW02).

Effects of the pal 4 DNA endonuclease activity on *E. coli*

We have previously observed (11), that the production in *E. coli* of the wild type version (pal 4/A) of a putative mitochondrial al 4 intronic translational product is highly toxic to *E. coli*. We noticed that the mutated form pal 4/B and pal 4/C can be produced without affecting bacterial growth. To further analyze the effects of the pal 4 DNA endonuclease on *E. coli*, we examined the structure of the *E. coli* chromosome after different forms of the pal 4 protein were produced. The effects of the DNA endonuclease activity on the *E. coli* chromosome were analyzed by pulsed field gel electrophoresis (Figure 5) of Not I digests. In a control experiment, where a truncated form of the pal 4 protein is produced, a stable pattern of the *E. coli* DNA can be

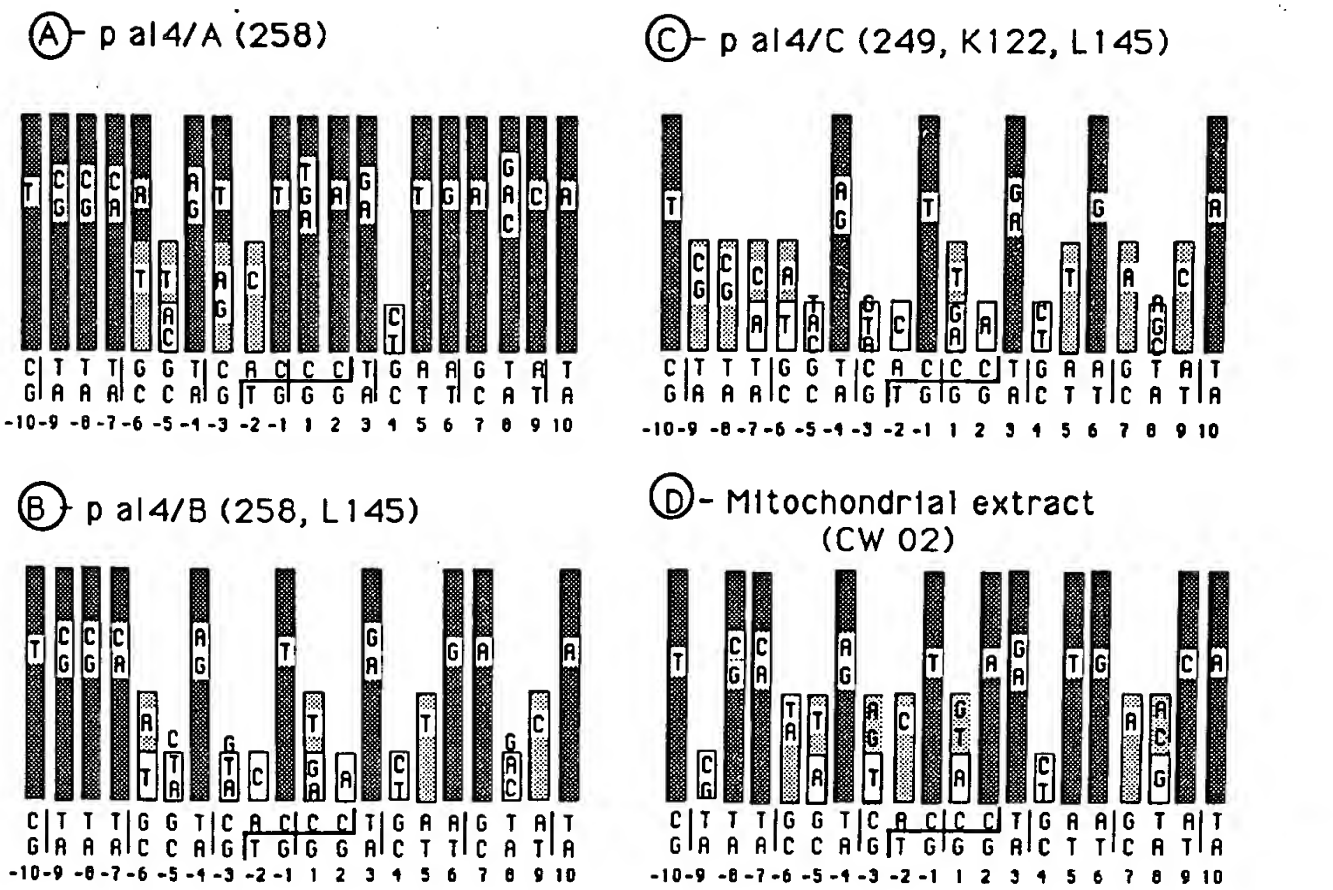


Figure 3. DNA endonucleolytic properties of the different forms of the pal 4 intronic protein on different single mutants of its A4-A5 target sequence. Long dark bars correspond to single mutants that were cut like the wild-type A4-A5 target sequence. Half-length grey bars correspond to mutants that affect the cleavage efficiency of the pal 4 DNA endonuclease. Short white bars correspond to single mutants that were not cleaved. All the indicated mutated sequences were from the upper strand of the target sequence. pal 4/A/B/C are different forms of the *E. coli*-made DNA endonuclease (see text) and the corresponding results have been obtained from in vivo experiments. Mitochondrial extract properties have been examined in vitro as indicated in Fig. 2 and in Materials and Methods.

observed which is quite similar to the pattern obtained when the pal 4/B protein is produced. In contrast, the production of the protein pal 4/A leads to dramatic effects on the *E. coli* chromosome structure. When the *E. coli* cells are left at 40°C for 10 minutes, specific cleavage products clearly appear (arrows in Fig. 5). Longer periods at 40°C lead to the complete degradation of the *E. coli* chromosome (data not shown).

DISCUSSION

In this work we have examined, *in vivo* or *in vitro*, the recognition site specificity of the intronic DNA endonuclease pal 4 (*I-SceII*) as a function of the different forms of the protein made in *E. coli* or from cellular extracts of a mitochondrial mutant. Surprisingly, we found that recognition of the pal 4 site is a complex phenomenon. The different 17 bp forms of the minimal recognition site do not have an equivalent role on the control of the DNA endonucleolytic activity. By taking advantage of the 35 single mutations that we constructed we observed that, in fact, many base changes do not affect cleavage *in vivo* or *in vitro*. It looks as if only a few base pairs scattered along the 17 bp minimal recognition site would be essential. Considering our results, a mean consensus sequence mainly based on the properties of the yeast mitochondrial extract (figure 3), can be written as follows:

5' N . (T) N N . G G N . C A N . C (C) N . G N N . (G) T N 3'

in which the indicated bases are important since at least one base change in these positions severely affects the cleavage efficiency (bases between brackets correspond to less critical positions). Interestingly, the *E. coli*-made protein pal 4/B behaves rather like the DNA endonuclease from mitochondrial extracts whereas its 'wild type' form, pal 4/A, has a very much reduced specificity. In the case of pal 4/A only 5 positions in the target sequence seem to be critical (G+4, A-2, C-3, G-5, G-6). The only difference between the two genes coding either for pal 4/A or pal 4/B is in the codon at position 145 which is either ACU or CUU. This indicates that when the CUU codon is translated as a Leu, which is at variance with the accepted view of the mitochondrial genetic code (33), the properties of the corresponding *E. coli*-made protein (pal 4/B) are rather similar to those of the mitochondrial protein accumulating in the yeast mutant CW02. This again (see (11)) raises some doubts on the meaning of the CUU codon in mitochondria.

Although it is tempting to speculate about the variations in the specificity of the different proteins we will only stress the fact that only very limited variations in the structure of the protein

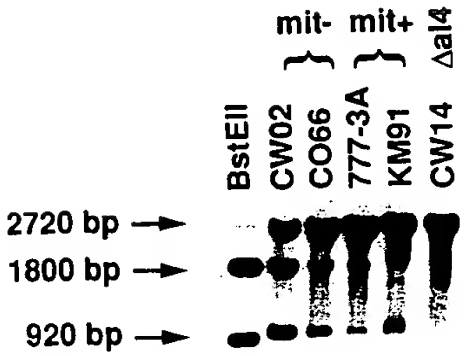


Figure 4. DNA endonuclease activity of the pal 4 protein (*I-SceII*) in mutant and wild-type yeast strains. The pUC A4-A5 plasmid (Fig.1) which was linearized at the ScaI site was incubated with mitochondrial extracts of the different strains or was digested with BstEII (left). Cleavage was detected by Southern blot analyses with the pUC plasmid as a probe. CWO2 and C066 (Gift from P. Netter, CGM Gif/Yvette) accumulate the al4 intron because of the absence of the bl4 RNA maturase or the presence of a mutation in the al4 intron. 777-3A and KM 91 are, respectively, haploid and diploid wild-type strains. CW14 is a control in which the al4 intron is deleted from the mitochondrial genome. The amount of extracted proteins used for each assay was: 11 μg (CW 02); 13 μg (CO 66); 14 μg (777-3A); 10μg (KM 91) and 15 μg (CW14).

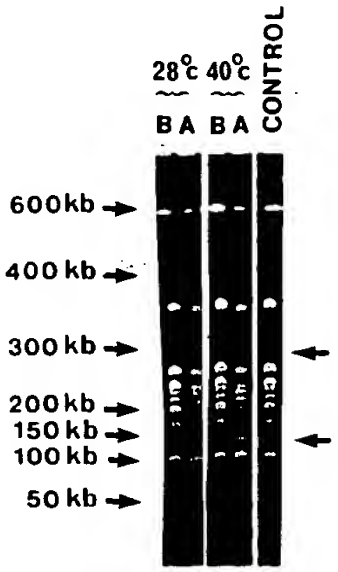


Figure 5. Ethidium-stained pulsed-field gel of NotI digests of *E. coli* DNA. NotI digestions were carried out on *E. coli* strains in which no foreign protein was expressed (control) or in strains in which p28 al4/B (lanes B) or p28al4/A (lanes A) were expressed at 40°C under the control of the lambda P_R promoter, which is almost completely blocked at 28°C (left).

pal 4 are able to induce differences in its ability to recognize and cut the different mutated sequences. In mitochondria the actual form of the pal 4 DNA endonuclease in the wild type strain is unknown. Even in the case of mutant strains where the splicing of the al 4 intron is blocked, we do not know whether the DNA endonuclease activity is associated with the long p56 chimeric protein corresponding to the translation of the upstream exons and intron sequence or associated with the p30 intronic proteolytic

A4-A5 al4 homing site	C	TTT	GGT	CAC	CCT	GAA	GTA
A4-al4 Exon-intron junction	C	TTT	GGT	CAA	ACA	GTG	GCC
B4-B5 bl4 flanking sequences	T	TTA	GGT	CAT	CCT	GAT	AAC
E.coli LACT	C	TTT	GGT	CAC	GAT	GAT	GTC
E.coli nrdB	C	TTT	GGT	CAG	CCG	GTC	AAC
Consensus sequence	N	(T)NN	GGN	CAN	C(G)N	(G)NN	GTN

Table 1. Sequences in the yeast mitochondrial genome and in the *E. coli* genome that are similar to the A4-A5 target sequence (upper line). Most of the essential bases (named) or important bases (bracketed) in the lower line are found in the 5' exon-intron junction of A4-al4 or in the exonic sequences flanking the fourth intron in the cytochrome b gene. A systematic perusal of the available *E. coli* sequences reveals two homologous sequences in LACT (43) and in the gene coding for the ribonucleoside diphosphate reductase (44).

product that we recently observed. Our observation (11) that the pal 4 (28 Kd) engineered intronic protein produced in *E. coli* is a DNA endonuclease would rather favor the last possibility but this does not exclude a possible role of the precursor in the modulation of the specificity of the endonuclease. The study of the wild-type DNA endonuclease activity is made extremely difficult because of the very low amounts of this protein in the wild-type cell where we cannot even immunologically detect it (20, 36). Only recently have we succeeded in observing a specific DNA endonuclease activity in different wild-type cells (Figure 4). Unfortunately this weak DNA endonucleolytic activity did not allow a systematic analysis of the recognition specificity. Purification of the wild-type protein, now in progress, will make this experiment feasible.

These site recognition studies allow us to explain why some forms of the pal 4 protein are highly toxic to *E. coli*. It has clearly been shown, with EcoRI enzyme, that multiple double-strand breaks in the *E. coli* chromosome cannot be repaired while a few such breaks are usually repaired with the help of DNA ligase (37). The expression of pal 4/A in *E. coli* clearly introduces several breaks in the *E. coli* chromosome (Figure 5) while pal 4/B has no evident effects. Thus we can suppose that either pal 4/A introduces too many breaks in the *E. coli* chromosome or it remains tightly bound to the ends of the DNA after cleavage and prevents the action of the DNA ligase, as in the case of the lambda terminase (38). This last hypothesis should be considered after the observation that DNA ends generated *in vivo* by the pal 4 DNA endonuclease are highly resistant to the action of DNA exonucleases (11). The fact remains, however, that pal 4/A recognizes and cuts DNA sequences in the *E. coli* chromosome (Figure 5). A survey of the available *E. coli* sequences reveals that two sequences are similar to the A4/A5 target sequence (Table 1). These two sequences are located in the unknown gene LACT(43) and in the gene coding for ribonucleoside reductase (44) which, in the phage T4, contains an intron (39).

The observed flexibility of the pal 4 DNA endonuclease specificity wich has to be confirmed *in vivo*, suggests that it could

provide a switch mechanism to control intron propagation. Variations in the specificity might depend on the presence of other cellular factors and this would imply that an intron might insert in locations different from its normal homing site, which suggests a real transposition of the intron. Such an ectopic integration of the intron could explain the presence of homologous introns in various genome locations (reviewed in (8, 40))

Moreover, the complexity of the recognition site was surprising since in the case of *I-SceI*, the first intronic DNA endonucleolytic activity which has been studied in detail, there was a minimal sequence degeneracy that was tolerated within the 18 base pair long recognition site (41). It should be noticed, however, that the pal 4 intronic DNA endonuclease recognizes and cuts a sequence coding for a protein. In that respect, it is interesting to note (Figure 3) that sequence degeneracy is maximal at each third base codon position; conversely, the critical bases are always found in the first or second codon positions. If confirmed in the case of the other intronic DNA endonucleases, the raison d·etre of such a property might be to favor the horizontal transfer of introns in sequences coding for homologous amino acid sequences but which differ in their third base positions. A survey of the different known intron locations (reviewed in (8, 40)) supports this possibility.

More generally, the observed flexibility in the recognition of A4-A5 exon sequences by the pal 4 DNA endonuclease is reminiscent of recent observations on the properties of the mitochondrial transcriptional machinery, which can accommodate several different promoter sequences (42). On the other hand, by its relative complexity, the recognition sequence of the pal 4 DNA endonuclease (*I-Sce II*) is similar to that of the HO nuclease which initiates mating-type interconversion in yeast (21).

ACKNOWLEDGEMENTS

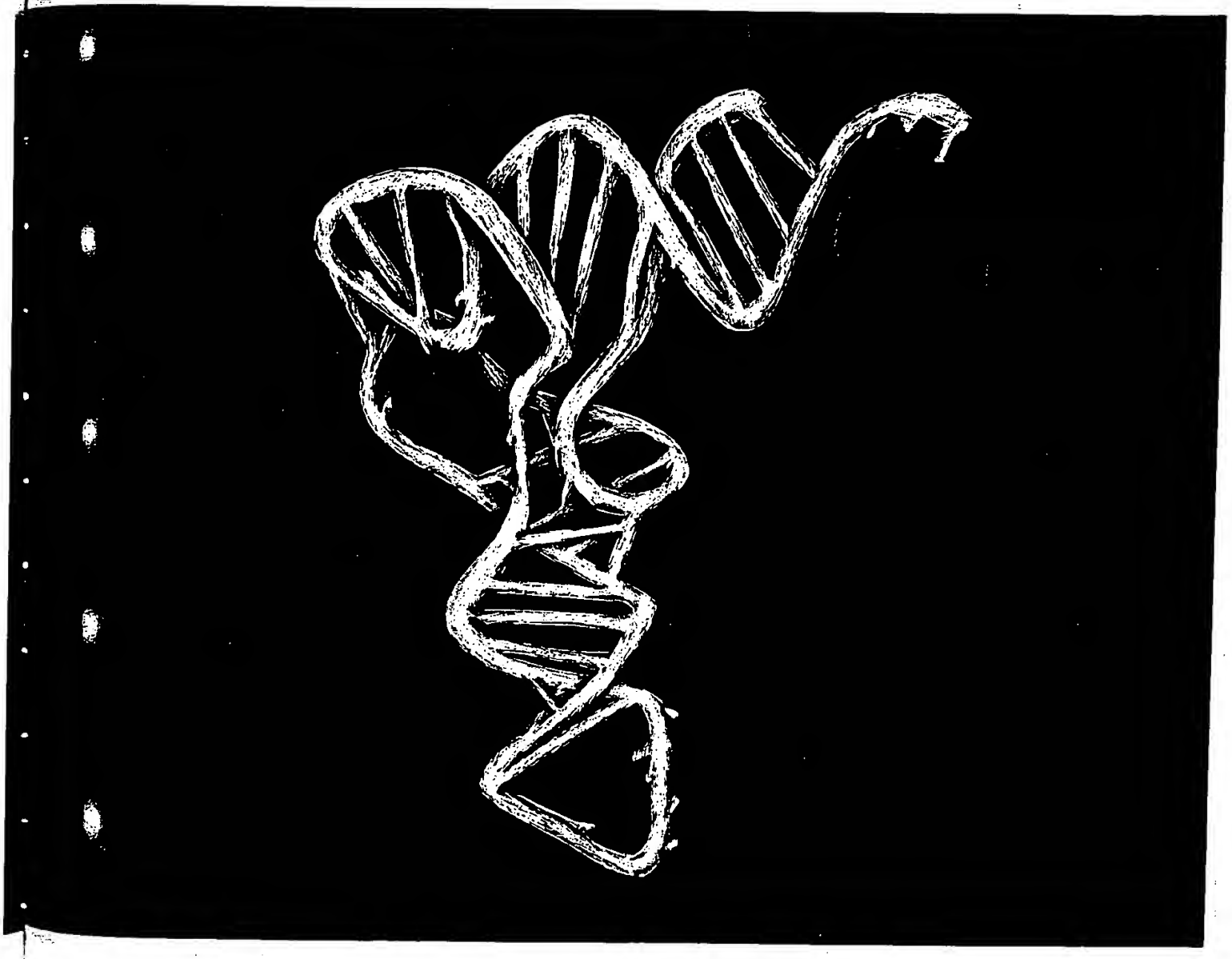
We thank Kyle Tanner and Josette Banroques for critical reading of the manuscript, Pierre Netter for generous gift of yeast mutants, Javier Perea and Thierry Soussi for helpful discussions, Francine Creusot for initiating some of the experiments and Francois Caron for his generous help in pulsed-field electrophoresis experiments. B.S is a recipient of an MRES fellowship. This work was supported by grants from CNRS (URA 1302), by INSERM (no. 891007) and by Université P. et M. Curie.

REFERENCES

1. Jacq C., Lazowska J. , Slonimski P.P. (1980) *C.R. Acad. Sc. Paris, sér. D* **290**, 89–92.
2. Colleaux L., d'Auriol L. , Betermier M., Cottarel G., Jacquier A., Galibert F., Dujon B. (1986) *Cell* **44**, 521–533.
3. Zinn A.R., Butow R. A. (1985) *Cell* **40**, 887–893.
4. Kotylak Z., Lazowska J. , Hawthorne D.C., Slonimski P.P. (1985) *Achievments and perspectives in mitochondrial research* **2**, 1–20.
5. Michel F., Lang B. F. (1985) *Nature* **316**, 641–643.
6. Lambowitz A.M. (1989) *Cell* **56**, 323–326.
7. Scazzochio C. (1989) *TIG* **5**, 168–172.
8. Dujon B. (1989) *Gene* **82**, 91–114.
9. Belfort M. (1989 Jul) *TIG* **5**, 209–213.
10. Perlman P.S., Butow R. A. (1989) *Science* **246**, 1106–1109.
11. Delahodde A., Goguel V. , Becam A.M., Creusot F., Perea J., Banroques J., Jacq C. (1989) *Cell* **56**, 431–441.
12. Wenzlau J.M., Saldanha R. J., Butow R.A., Perlman P.S. (1989) *Cell* **56**, 421–430.
13. Quirk S.M., Bell-Pedersen D. , Belfort M. (1989) *Cell* **56**, 455–465.
14. Chu, F. K., Maley, G., Pedersenlane, J., Wang, A. M. & Maley, F. (1990 May) *Proc. Natl. Acad. Sci. USA* **87**, 3574–3578.
15. Meunier B., Tian G. L., Macadre C., Slonimski P.P., Lazowska J. (1990) *Structure , function and biogenesis of energy transfer systems, E.Quagliarello, S. Papa, F. Palmieri, C.Saccone eds Elsevier* 169–174.
16. Michel F. (1984) *Curr. Gen.* 307–317.
17. Labouesse M., Netter P. , Schroeder R. (1984) *Eur. J. Biochem.* **144**, 85–93.
18. Banroques J., Delahodde A. , Jacq C. (1986) *Cell* **46**, 837–844.
19. Banroques J., Perea J. , Jacq C. (1987) *EMBO J.* **6**, 1085–1091.
20. Jacq C., Banroques J. , Becam A.M., Slonimski P. Guiso N., Danchin A. (1984) *EMBO J.* 3, 1567–1572.
21. Nickoloff J.A., Jeffrey D. S., Heffron F. (1990) *Mol.Cel.Biol.* 1174–1179.
22. Flamm E.L., Weisberg J. (1985) *J. Mol. Biol.* **183**, 117–128.
23. Kunkel T.A., Roberts J. D., Zakour R.A. (1987) *Methods Enzymol.* **154**, 367–382.
24. Carter P., Bedouelle H. , Winter G. (1985) *Nucl. Acids Res.* **13**, 4431–4443.
25. Kotylak Z., Slonimski P. P. (1977) *Mitochondria 1977* 83–89.
26. Schroeder R., Slonimski P. P. (1983) *Mitochondria* 203–209.
27. DelSal G., Manfioletti G. , Schneider C. (1988) *Nucl. Acids. Res.* **16**, 9878.
28. Sambrook J., Fritsch E. F., Maniatis T. (1989) *Molecular cloning : a laboratory manual, 2nd ed.*
29. Sanger F., Nicklens S. , Coulson A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
30. Church G.M., Gilbert W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1991–1996.
31. Huspedth M., Shumard D. S., Tatti K.M., Grossman L.I. (1980) *Biochem. Biophys. Acta* **610**, 221–228.
32. Chu et al. (1986) *Science* **234**,
33. Bonitz S.G., Berlani R. , Coruzzi G., Li M., Macino G., Nobrega F.G., Nobrega M.P., Thalenberg B.E., Tzagaloff A. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 3167–3170.
34. Dujardin G., Jacq C. , Slonimski P.P. (1982) *Nature* **298**, 628–632.
35. Labouesse M., Slonimski P. P. (1983) *EMBO J.* **2**, 269–276.
36. Goguel V., Perea J. , Jacq C. (1989) *Curr. Genet.* **16**, 241–246.
37. Heitman T., Zinder N. D., Model P. (1989) *Proc.Natl.Acad. Sci.USA* **86**, 2281–2285.
38. Roberts D., Kleckner N. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6037–6041.
39. Gott J.M., Shub D. A., Belfort M. (1986) *Cell* **47**, 81–87.
40. Wolf K., Giudice L. Del (1988) *Advances in Genetics* **25**, 185–308.
41. Colleaux L., d'Auriol L. , Galibert F., Dujon B. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6022–6026.
42. Fisher R.P., Parisi M. A., Clayton D.A. (1989) *Genes Dev.* 2202–2217.
43. Buvinger W.E., Lampel K.A., Bojanowski R.J., Riley M. (1984) *J. Bacteriol.* 159, 618–623.
44. Carlson J., Fuchs J.A., Messing J. (1984) *Proc. Natl. Acad. Sci.* 81, 4294–4297.

VOLUME 18 NUMBER 19 OCTOBER 11, 1990

Nucleic Acids Research



IRL PRESS
at
OXFORD UNIVERSITY PRESS
Oxford New York Tokyo

ISSN 0305 1048 Coden NARHAD